The Korean Communications
in Statistics Vol. 9, No. 3, 2002
pp. 825-834

Posterior Consistency of Bayesian Inference of Poisson Processes[1]

Yongdai Kim[2]

Abstract

Poisson processes are widely used in reliability and survival analysis. In particular, multiple event time data in survival analysis are routinely analyzed by use of Poisson processes. In this paper, we consider large sample properties of nonparametric Bayesian models for Poisson processes. We prove that the posterior distribution of the cumulative intensity function of Poisson processes is consistent under regularity condtions on priors which are Levy processes.

Keywords : Poisson process, Levy process, posterior consistency

1. Introduction

Interest in statistics inference for Poisson process has grown rapidly in the reliability engineering and survival analysis ever since Ascher (1968) proposed the use of the Poisson process as a realistic tool of modelling more complex reliable system suitable for statistical analysis. The parameter which characterizes a Poisson process is the cumulative intensity (or the derivative of it called intensity rate). Early approaches are parametric in nature, and it is assumed that the cumulative intensity, or the intensity rate, is of some linear or parametric form (Brown 1972, Lewis 1972, Snyder 1975). A Bayesian parametric approach was considered by Clevenson and Zidek (1977) who assumed a linear intensity rate, and by Grenander (1981) who assumed that the cumulative intensity is a step function. The Bayesian nonparametric approach was discussed by Lo (1982) and Kim (1999).

In contrast to the construction of suitable priors and their computation, theoretical properties of the posterior distribution of the cumulative intensity have received relatively little attention. Since Freedman and Diaconis (1986) described a possibility of posterior inconsistency when the parameter space is large, many researches on the issue of posterior consistency have been done

1) This work is supported (in part) by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.
2) Assistant Professor, Department of Statistics, Ewha Womans University, Seoul, 120-750, Korea.
E-mail : ydkim@mm.ewha.ac.kr

(Barron 1988, Barron et al. 1997, Ghoshal et al. 1998, Shen and Wasserman 1998, Kim and Lee 2001). However, posterior consistency of the Poisson process model has not been considered yet. In this paper, we consider the issue of posterior consistency for the Poisson process with Levy processes as priors. We prove that under certain conditions the posterior distribution is consistent.

The paper is organized as follows. In section 2, the model is stated and theories on Levy processes are reviewed. In section 3, the main theorem for the posterior consistency is given with regularity conditions on priors and its proof is given in section 4.

2. The model and Levy processes

A stochastic process $N=\{N(t): t\in[0,\infty)\}$ is called a Poisson process with cumulative intensity Λ (a nondecreasing function with $\Lambda(0)=0$) if N is a counting process and the process $M=\{M(t)=N(t)-\Lambda(t), t\in[0,\infty)\}$ is a martingale with respect to $\{\mathfrak{I}_t\}$ where $\mathfrak{I}_t$ is a sigma filed generated by $\{N(s), s\in[0,t]\}$.

Remark. The standard definition of the Poisson process is that $N(.)$ is an independent increment process and for each t, $N(t)$ is a Poisson random variable with mean $\Lambda(t)$. If Λ is continuous, our definition is equivalent to the standard definition, but they are not the same otherwise (see Jacod and Shiryaev 1987 for details).

Suppose $N_1,\ldots,N_n$ are i.i.d. Poisson process with cumulative intensity Λ. For prior distribution of Λ, we consider a Levy process. As is well known, any nondecreasing Levy process is a sum of a deterministic function and a jump process, and we assume that the deterministic function vanishes elsewhere. Note, that most of practically important processes such as beta process (Hjort 1990) and gamma process (Lo 1982) are such Levy process. In the remaining of this section, we review basic facts of Levy processes as priors of Λ.

Note that not all Levy process can be prior distribution of Λ since $\Delta\Lambda(t)$ should be bounded by 1. This is because the Poisson process is defined by the martingale argument (see Andersen et al. 1996 for details). In spite of this disadvantage, we use the definition of the Poisson process via the martingale argument since the posterior distribution is given in a very nice form, which makes it possible to study large sample properties of the posterior distribution.

Kim (1999) uses the following characterization of Levy processes whose jump sizes are bounded by 1. For any given Levy process $\Lambda(t)$ on $[0,\infty)$ with $0\le\Delta\Lambda(t)\le 1$, there exists a unique random measure μ on $[0,\infty)\times[0,1]$ such that

$$\Lambda(t) = \int_{[0,t]\times[0,1]} x\mu(ds, dx). \tag{1}$$

In fact, μ is defined by $\mu([0,t]\times[0,u]) = \sum_{s\le t} I(\Delta\Lambda(s) \le u)$ for all $t\in[0,\infty)$ and $u\in[0,1]$. Since μ is a Poisson random measure, there exists a unique σ-finite random measure ν on $[0,\infty)\times[0,1]$ such that

$$E(\mu([0,t]\times[0,u])) = \nu([0,t]\times[0,u]). \tag{2}$$

Conversely, for a given σ-finite measure ν satisfying $\int_{[0,t]\times[0,1]} \nu(ds, dx) < \infty$ for all $t>0$, there exists a unique Poisson random measure which satisfies (2) and so we can construct a Levy process Λ through (1). Conclusively, we can use ν to characterize a Levy process Λ. We call ν the Levy measure of Λ. Details about Levy proceses and Poisson measure can be found in Kim (1999) and Kim and Lee (2001).

For a given Levy process Λ with Levy measure ν, we can easily calculate mean and variance by the following equations (Kim 1999):

$$E(\Lambda(t)) = \int_0^t \int_0^1 x\nu(ds, dx) \tag{3}$$

and

$$Var(A(t)) = \int_0^t \int_0^1 x^2\nu(ds, dx) - \sum_{s\le t}\left(\int_0^1 x\nu(s, dx)\right)^2 \tag{4}$$

For a prior distribution of the cumulative intensity Λ, we assume that Λ is a Levy process with Levy measure ν. Further, we suppose that ν is given by

$$\nu([0,t]\times[0,u]) = \int_0^t \int_0^u f_s(x)\,dx\,ds. \tag{5}$$

Then the posterior distribution of Λ given $(N_1, \ldots, N_n)$ is again a Levy process with Levy measure ν^p given by

$$\begin{aligned} \nu^p([0,t]\times[0,u]) &= \int_0^t \int_0^u (1-x)^n f_s(x)\,dx\,ds \\ &+ \int_0^t c_n(s)^{-1} \int_0^u x^{\Delta N_.(s)}(1-x)^{n-\Delta N_.(s)} f_s(s)\,ds \frac{1}{\Delta N_.(s)}\, dN_.(s) \end{aligned} \tag{6}$$

where $N_. = \sum_{i=1}^{n} N_i$ and

$$c_n(s)=\int_0^1 x^{\Delta N.(s)}(1-x)^{n-\Delta N.(s)} f_s(x)\,dx.$$

The proof is in Kim (1999).

3. Main results

In this section, we give sufficient conditions for the posterior consistency of Λ when the prior is a Levy process. Posterior consistency means that the posterior distribution of Λ converges weakly to the point mass at Λ^* on $D[0,\tau]$ for all $\tau>0$ if $N_1,\ldots,N_n$ are i.i.d. Poisson processes with the common cumulative intensity Λ^*. Here, $D[0,\tau]$ is the set of right continuous functions with left limit existing equipped with the uniform metric. Assume that $f_s(x)$ in (5) is given by $f_s(x)=\frac{1}{x}g_s(x)\lambda_0(s)$ where $\int_0^1 g_t(x)dx=1$ for all $t>0$. Assume that $\lambda_0(t)$ is bounded and positive on $(0,\infty)$.

For the posterior consistency, we need the following two conditions:

(C1) $\sup_{t\in[0,\infty),x\in[0,1]}(1-x)g_t(x)<\infty$

(C2) there exists a function $h(t)$ such that $\lim_{x\to 0}\sup_{t\in[0,\infty)}|g_t(x)-h(t)|=0$ and $0<\inf_{t\in[0,\infty)}h(t)\le\sup_{t\in[0,\infty)}h(t)<\infty$.

Theorem 1. *Suppose Λ^* is continuous. Under (C1) and (C2), the posterior distribution of Λ given $(N_1,\ldots,N_n)$ is consistent.*

The conditions (C1) and (C2) are used by Kim and Lee (2001) for proving the posterior consistency of the survival function with right censored data. Also, they show that beta process and gamma process satisfy (C1) and (C2) and hence yield consistent posteriors. In this view, this paper extends Kim and Lee's result to Poisson process models.

4. Proofs of Theorem 1

Lemma 1. *For $i=0, 1, \cdots$, $s\in[0,\ \tau]$, $c>0$ and $\lambda>0$,*

$$\limsup_{n\to\infty}\ \sup_{s\in[0,\tau]} n^{i+1}\int_{\lambda/n}^1 x^i(1-x)^{n-c}dx\le\sum_{k=0}^{i}C_k\lambda^k\exp(-\lambda)\ a.s.,$$

where C_k are positive constants, for $k=0, 1, \cdots, i$, depending on i and c, but not on λ and s.

Proof. Without loss of generality, assume $0<\lambda/n<1$ and $n-c>0$. We prove this lemma by induction on i. First, suppose $i=0$. Since

$$n\int_{\lambda/n}^{1}(1-x)^{n-c}dx=\frac{n}{n-c+1}(1-\frac{\lambda}{n})^{n-c+1},$$

the result follows for $i=0$.

Suppose the result holds for $i-1$ with i=1, 2, $\cdots$. Then, using integration by part,

$$n^{i+1}\int_{\lambda/n}^{1}x^{i}(1-x)^{n-c}dx \leq \frac{n^{i+1}}{n-c+1}(\frac{\lambda}{n})^{i}(1-\frac{\lambda}{n})^{n-c+1}$$

$$+\frac{ni}{n-c+1}n^{i}\int_{\lambda/n}^{1}x^{i-1}(1-x)^{n-c+1}dx.$$

Using the induction assumption, the result follows for all $i=0, 1, \cdots$. □

Lemma 2. *Suppose (C1) holds. Then, for i=0, 1, $\cdots$,*

$$\sup_{t\in[0,\tau]}\int_{0}^{t}\int_{0}^{1}x^{i}(1-x)^{n}g_{s}(x)\,dx\,\lambda_{0}(s)\,ds\rightarrow 0,\ \textit{with probability } 1.$$

Proof. Let $g^{*}=\sup_{x\in[0,1],s\in[0,\tau]}(1-x)g_{s}(x)$, which is finite by (C1). Using the fact $|x|^{i}\leq 1$ for all $i=0, 1,\cdots$, we have

$$\int_{0}^{t}\int_{0}^{1}x^{i}(1-x)^{n}g_{s}(x)\,dx\,\lambda_{0}(s)\,ds$$

$$\leq\int_{0}^{\tau}\int_{0}^{1}x^{i}(1-x)^{n}g_{s}(x)\,dx\,\lambda_{0}(s)\,ds$$

$$\leq\int_{0}^{\tau}\int_{0}^{1}(1-x)^{n-1}(1-x)\,g_{s}(x)\,dx\,\lambda_{0}(s)\,ds$$

$$\leq\int_{0}^{\tau}\frac{g^{*}}{n}\lambda_{0}(s)\,ds$$

$$\leq\frac{g^{*}\Lambda_{0}(\tau)}{n}\rightarrow 0$$

where $\Lambda_{0}(\tau)=\int_{0}^{\tau}\lambda_{0}(s)ds$, and the proof is completed. □

Lemma 3. *Suppose (C1) and (C2) hold. Then, for i=0, 1, $\cdots$,*

$$\sup_{s\in[0,\tau]}\int_{0}^{1}x^{i}(1-x)^{n-1}|g_{s}(x)-h(s)|\,dx=o(\frac{1}{n^{i+1}}),\ \textit{with probability } 1.$$

Proof. Let

$$M=\sup_{x\in[0,1],s\in[0,\tau]}(1-x)[g_{s}(x)+h(s)].$$

Note that $M<\infty$ by (C1) and (C2). Choose an arbitrary large positive number λ. For all positive integer n such that $\lambda/n < 1$, define

$$d_n = \sup_{x\in[0,\lambda/n],s\in[0,\tau]} |g_s(x)-h(s)|.$$

By (C2), $\lim_{n\to\infty} d_n = 0$. For all n with $\lambda/n<1$ and $s \in [0, \tau]$, we have

$$\begin{aligned}
&n^{i+1}\int_0^1 x^i(1-x)^{n-1}|g_s(x)-h(s)|\,dx \\
&\le n^{i+1}\left[\int_0^{\lambda/n}+\int_{\lambda/n}^1\right]x^i(1-x)^{n-1}|g_s(x)-h(s)|\,dx \\
&\le n^{i+1}d_n\int_0^{\lambda/n}x^i(1-x)^{n-1}dx+n^{i+1}\int_{\lambda/n}^1 x^i(1-x)^{n-2}(1-x)|g_s(x)-h(s)|\,dx \\
&\le n^{i+1}d_n\left(\frac{\lambda}{n}\right)^i\int_0^{\lambda/n}dx+n^{i+1}M\int_{\lambda/n}^1 x^i(1-x)^{n-2}dx \\
&\le \lambda^{i+1}d_n+n^{i+1}M\int_{\lambda/n}^1 x^i(1-x)^{n-2}dx.
\end{aligned}$$

By lemma 1 and the fact that $\lim_{n\to\infty} d_n = 0$,

$$\limsup_{n\to\infty}\ \sup_{s\in[0,\tau]} n^{i+1}\int_0^1 x^i(1-x)^{n-1}|g_s(x)-h(s)|\,dx \le \sum_{k=0}^{i} C_k\lambda^k\exp(-\lambda),$$

for some positive constants C_k independent of λ, for $k=0, 1, 2, \cdots, i$. Since λ is arbitrary large, the result follows. □

Lemma 4. *Suppose* (C1) *and* (C2) *hold. Then , for* $i=0, 1, \cdots$,

$$\sup_{s\in[0,\tau],\triangle N_.(s)=1}\left|\int_0^1 x^i(1-x)^{n-1}\left(\frac{g_s(x)}{k_n(s)}-n\right)dx\right| = o\left(\frac{1}{n^i}\right)$$

where

$$k_n(s)=\int_0^1(1-x)^{n-1}g_s(x)\,dx.$$

Proof. For $s\in[0,\tau]$ with $\triangle N_n(s)=1$,

$$\begin{aligned}
&\left|\int_0^1 x^i(1-x)^{n-1}\left(\frac{g_s(x)}{k_n(s)}-n\right)dx\right| \\
&\qquad\le \frac{1}{k_n(s)}\int_0^1 x^i(1-x)^{n-1}\,|\,g_s(x)-h(s)\,|\,dx \\
&\qquad+\frac{1}{k_n(s)}\int_0^1 x^i(1-x)^{n-1}\,|\,h(s)-k_n(s)n\,|\,dx. \qquad (7)
\end{aligned}$$

It suffices to show supremums of two terms on the right hand side of (7) over all $s\in[0,\tau]$ with $\triangle N_.(s)=1$ converge to 0 with probability 1. In this proof, sup and inf are the supremum and infimum over all $s\in[o,\tau]$ with $\triangle N_.(s)=1$, respectively.

First, we have

$$\sup \mid nk_n(s) - h(s) \mid = \sup \mid n\int_0^1 (1-x)^{n-1} g_s(x)\, dx - n\int_0^1 (1-x)^{n-1} h(s) dx \mid$$

$$\leq \sup n\int_0^1 (1-x)^{n-1} \mid g_s(x) - h(s) \mid dx$$

$$= o(1) \text{ with probability 1,} \tag{8}$$

where the last equality is due to Lemma 3.

Consider the first term in (7). We have

$$\inf nk_n(s) \geq \inf h(s) - \sup \mid h(s) - k_n(s) n \mid \tag{9}$$

By (C2) and (8) we see that $nk_n(s) > 0$ all but finitely many n. Hence, lemma 3 implies

$$\sup \frac{n^{i+1}}{nk_n(s)} \int_0^1 x^i (1-x)^{n-1} \mid g_s(x) - h(s) \mid dx \to 0 \text{ , with probability 1.}$$

Finally, consider the second term in (7).

$$n^i \sup \frac{1}{k_n(s)} \int_0^1 x^i (1-x)^{n-1} \mid h(s) - k_n(s) n \mid dx$$

$$\leq n^i \sup \frac{h(s) - k_n(s) n}{k_n(s)} \int_0^1 x^i (1-x)^{n-1} dx$$

$$\leq \sup \frac{\mid h(s) - k_n(s) n \mid}{k_n(s) n} \frac{n^i \Gamma(i+1)}{(n+1) \cdots (n+i)}$$

Again using (C2) and (8) together we have

$$\inf k_n(s) n \geq \inf h(s) - \sup \mid h(s) - k_n(s) n \mid > 0$$

all but finitely many n with probability one. Hence the second term in (7) converges to 0 with probability due to (8). This completes the proof. □

Proof of Theorem 1. Since the posterior distribution of Λ is also a Lévy process and so it has nondecreasing sample paths with probability 1, it suffices to show that

$$E(\Lambda(t) \mid N_1, \ldots, N_n) \to \Lambda^*(t) \tag{10}$$

and

$$\mathrm{Var}\ (\Lambda(t) \mid N_1, \ldots, N_n) \to 0 \tag{11}$$

with probability one, for all $t \in [0, \tau]$.

Nothing that $k_n(s) = c_n(s)/\lambda_0(s)$ for $s > 0$ with $\triangle N_{.}(s) = 1$ and $N_{.}(0) = 0$ with probability one, we have

$$E(\Lambda(t) \mid N_1, \ldots, N_n) = \int_0^t \int_0^1 (1-x)^n g_s(x)\, dx\, \lambda_0(s)\, ds + \int_0^t \frac{1}{k_n(s)} \int_0^1 x(1-x)^{n-1} g_s(x)\, dx\, dN_.(s) \qquad (12)$$

By Lemma 2, the first term on the right hand side of (12) converges to 0 with probability1. By adding and subtracting the same quantity and using Lemma 4, the second term on the right hand side of (12) is rewritten by

$$\int_0^t n \int_0^1 x(1-x)^{n-1} \left| \frac{g_s(x)}{k_n(s)} - n \right| dx \frac{1}{n}\, dN_.(s) + \int_0^t n^2 \int_0^1 x(1-x)^{n-1} dx \frac{1}{n}\, dN_.(s) \qquad (13)$$

Note that

$$\sup_{t \in [0,\tau]} \left| \frac{N_.(t)}{n} - \Lambda_0(t) \right| \to 0 \qquad (14)$$

with probability 1 (see Andersen et al. 1996 for the proof). Hence, by (14) and Lemmas 4, the first term of (12) converges to 0 with probability 1. By the beta integral, the second term of (12) is

$$\int_0^t \frac{n}{n+1} \cdot \frac{1}{n}\, dN_.(s).$$

Since $\frac{n}{n+1} - 1 \to 0$, the second term of (12) converges to $\Lambda^*(t)$ by (14).

For (11) , we have

$$Var(\Lambda(t) \mid N_1, \ldots, N_n) = \int_0^t \int_0^1 x(1-x)^n g_s(x)\, dx\, \lambda_0(s)\, ds + \int_0^t \left[\frac{1}{k_n(s)} \int_0^1 x^2(1-x)^{n-1} g_s(x)\, dx - \left(\frac{1}{k_n(s)} \int_0^1 x(1-x)^{n-1} g_s(x)\, dx \right)^2 \right] dN_.(s) \qquad (15)$$

The first then on the right hand side of (15) converges to 0 with probability 1 by lemma 2. Since the integrand of the second term of (15) is the variance of the density

$$\frac{1}{k_n(s)} (1-x)^{n-1} g_s(x)\ , \text{ for } 0 < x < 1\ ,$$

it is less than or equal to

$$\frac{1}{k_n(s)} \int_0^1 x^2(1-x)^{n-1} g_s(x)\, dx\ .$$

Hence, the second term on the right hand side of (15) is less than or equal to

$$\int_0^t n \int_0^1 x^2(1-x)^{n-1} \left| \frac{g_s(x)}{k_n(s)} - n \right| dx \frac{1}{n}\, dN_{.}(s)$$
$$+ \int_0^t n^2 \int_0^1 x^2(1-x)^{n-1}\, dx \frac{1}{n}\, dN_{.}(s)$$
$$\le \int_0^t n \int_0^1 x^2(1-x)^{n-1} \left| \frac{g_s(x)}{k_n(s)} - n \right| dx \frac{1}{n}\, dN_{.}(s)$$
$$+ \int_0^t \frac{n\Gamma(3)}{(n+1)(n+2)} \frac{1}{n}\, dN_{.}(s)$$

By (14) and lemmas 4, it converges to 0 with probability 1. □

References

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical models based on counting processes*, Springer, NewYork.

[2] Ascher, H. (1984). Evaluation of repairable system reliability using the "Bad-as-old" concept. *IEEE Transactions on reliability*, **17**, 103-110.

[3] Barron, A.R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical Report, Univ. Illinois.

[4] Barron, A.R., Schervish, M.J. and Wasserman, L. (1999). The consystency of posterior distributions in nonparametric problems. *Ann. Statist.*, **27**, 5360561.

[5] Brown, M. (1972). Statistical analysis of nonhomogeneous Poisson processes. *Stochastic Point Process*. Edited by P.A.W. Lewis, 67-89, Wiley, New York.

[6] Clevenson, M.I. and Zidek, J.V. (1977). Bayes linear estimators of the intensity function of the nonstationary Poisson process. *J. Am. Statist. Assoc.*, **72**, 112-120.

[7] Ghoshal, S., Ghosh, J.K. and Ramamoorthi, R.V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.*, **27**, 143-158.

[8] Grenander, U. (1981). *Abstract Inference*, Wiley, New York.

[9] Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, **18**, 1259-1294.

[10] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*. Springer, New York.

[11] Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.*, **27**, 562-588.

[12] Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Ann. Statist.*, **29**, 666-686.

[13] Lewis, P.A.W. (1972). Recent results in the statistical analysis of univariate point processes. *Stochastic Point Process*. Edited by P.A.W. Lewis, 67-89, Wiley, New

York.
[14] Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. verw. Gebiete*, **59**, 55-66.
[15] Shen, X. and Wasserman, L. (1998). Rates of convergence of posterior distributions. *Ann. Statist.* **29**, 689-714.
[16] Snyder, D.L. (1975). *Random Point Processes*, Wiley, New York.

[2002년 7월 접수, 2002년 9월 채택]